Current & Future Technical Tools Survey For Development & Database Usage With Respondent Demographics

Craig Rupp

11/30/2022

IBM Developer

SKILLS NETWORK

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts (end of presentation)
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusions
- Appendix

EXECUTIVE SUMMARY

- The following slides will present survey findings from users on Stack Overflow regarding the following topics
  - Data from the survey polled users on current **programming languages** and **databases** used as well as future technologies that they would like to learn in each subject.
  - The survey also polled similarly for current & future usage for **platforms** leveraged for development, hosting, and services (ex. Heroku, AWS, Docker) as well as **web/programming frameworks** such as (ex. Django, React, Laravel).
  - Lastly, the respondent data is aggregated demographically by **region**, **age**, **gender**, and **education level.**
- Visualizations such as **bar graphs**, **tree charts**, **geo/map chart** among others are used to display top technologies used by our respondents

INTRODUCTION

- Analysis Overview
  - The following visualizations and data findings further seen in this presentation are categorical in nature by identifying the most common technologies used by our respondents with respective count aggregation for our grouped by categories such as how many users currently use the JS React Framework.
- Research Objectives
  - Highlight the top programming languages and databases used by developers currently as well as their desires for further expansion into other languages or databases.
  - Capture the array of web frameworks and platforms used by our respondents.
  - Identify demographics of survey respondents to associate language or database usage with our respondents make-up. (Ex : Gender, Education, etc)
- Points of Interest
  - What are the most common programming languages and database tools used by programmers today?
  - Do current usage trends match with future trends our respondents look to add to their skillset?
  - Do our respondents have a potential trend in where they are located, how much education they have received or their age?

METHODOLOGY

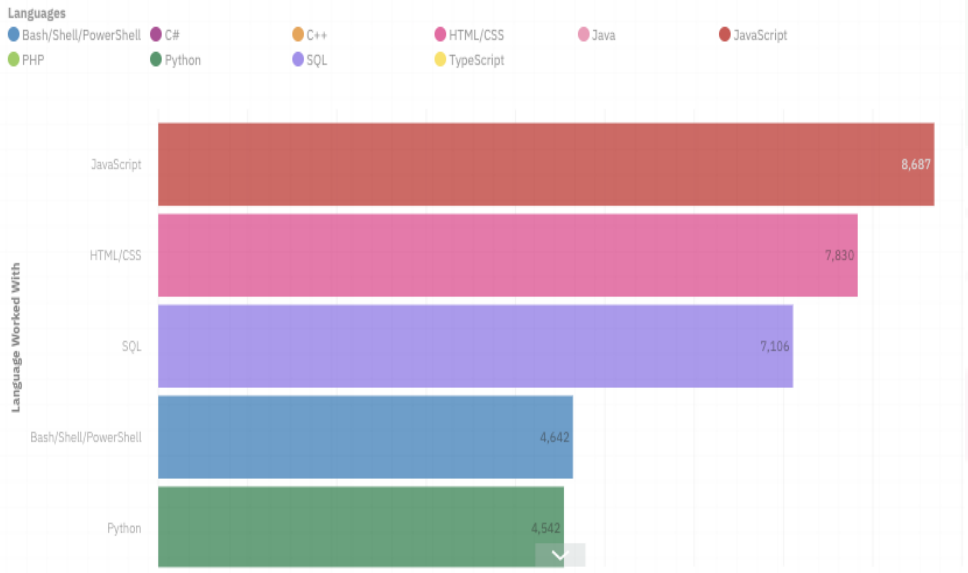
- Data Sources / Dashboard Tools
  - The Data Source used is a CSV file aggregating our survey respondent counts according to categories alluded to earlier in the presentation.
  - IBM's Cognos Dashboard was used to leverage the flat CSV file and create such visualizations to capture the counts of our programming and database listed options.
- Collected Data Objectives
  - Group categorical counts of survey respondents by technology to identify top current and future trends
  - Leverage visualizations to show pronounced trends for areas of interest
    - Ex – Hierarchical Bubble Chart to identify the most commonly used web frameworks
  - Discover **Demographic** related insights to highlight **where** our respondents live, **what** age trend we see for a majority of our respondents, and if a particular **education level** was most commonly attained by our respondents.

RESULTS GENERATION

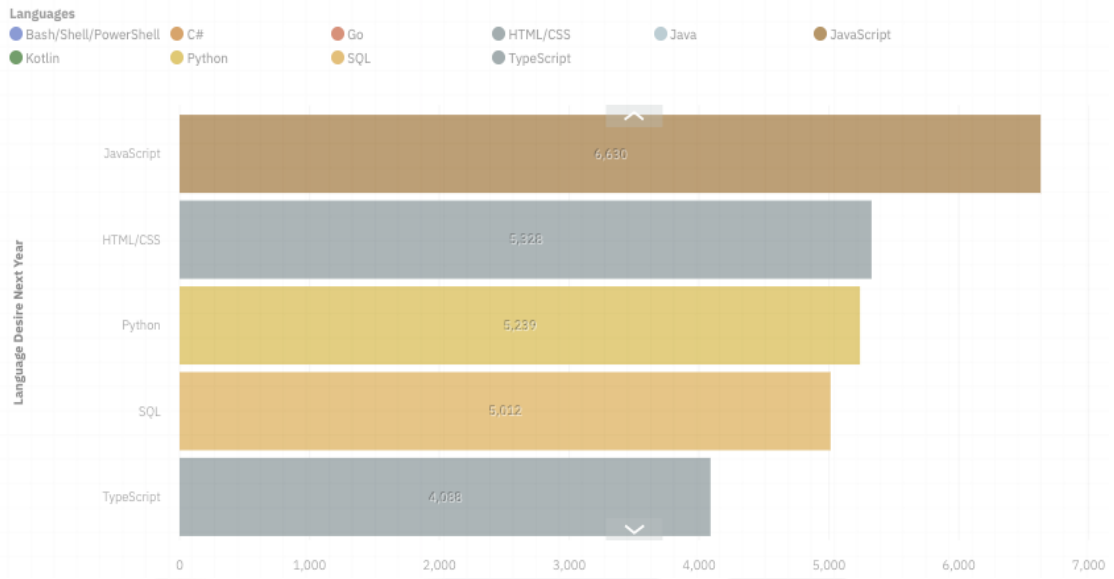
- Data Organization
  - The resulting visualizations are ordered in a Cognos aggregated Dashboard with three respective "tabs"
    - Current Technology Usage
    - Future Technology Usage
    - Respondent Demographics
  - The first two tabs focus on the Top 10 categorical choices for current/future technology usage while the platform choices are not limited to ten results.
- Data Analysis
  - Similar to either a SQL query or usage of a tabular data technology (Pandas/Excel), respective counts for each category is grouped by respective area of interest and limited to a select "highest" count to show relevant trends for our areas of analysis.
  - Grouped by counts were then organized into categorical count* type charts (bar, bubble, tree) to illustrate our top respondent areas.
  - Demographic visualizations (line graph, geo chart, pie chart) to see the percentage of our respondents demographic traits.

PROGRAMMING LANGUAGE TRENDS

Current Year



Next Year - Future



Top 5 Programming Languages

PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

Findings

* Below are a Few noteworthy findings

- **JavaScript** & **HTML** maintain top and second spot respectively
- New Entry "**TypeScript**" for a future trending language
- Higher **Python** Usage in subsequent years likely
- **SQL** still a bedrock for utility of uses across different database platform (Ex. - PostgreSQL, MySQL)
- Less of a Focus on "shell" scripting languages (Ex. - **Bash**, **PowerShell**)

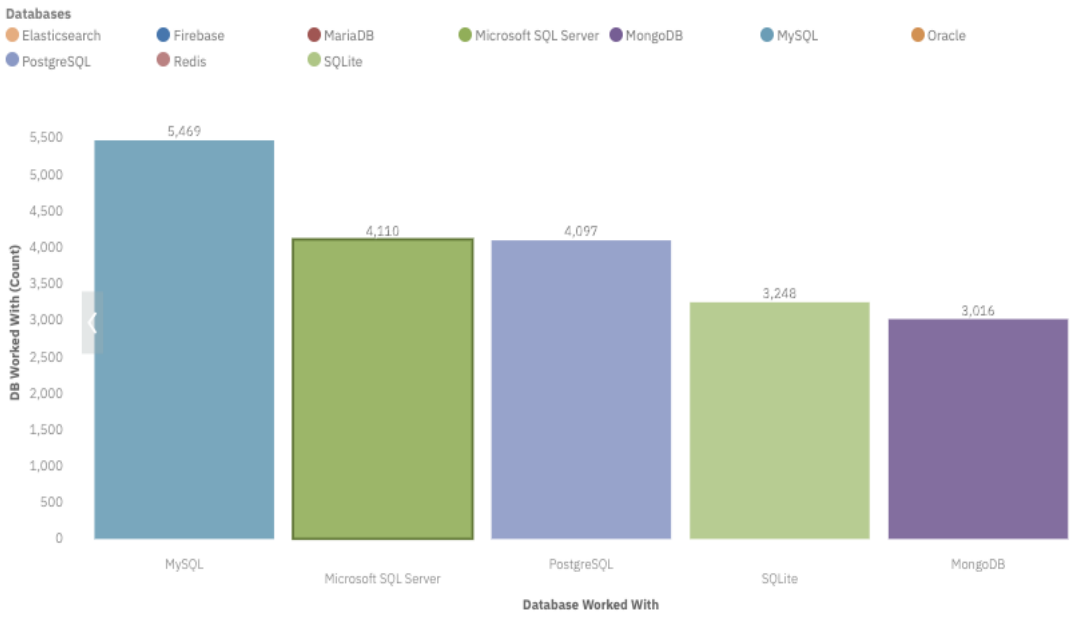
Implications

* And potential implications

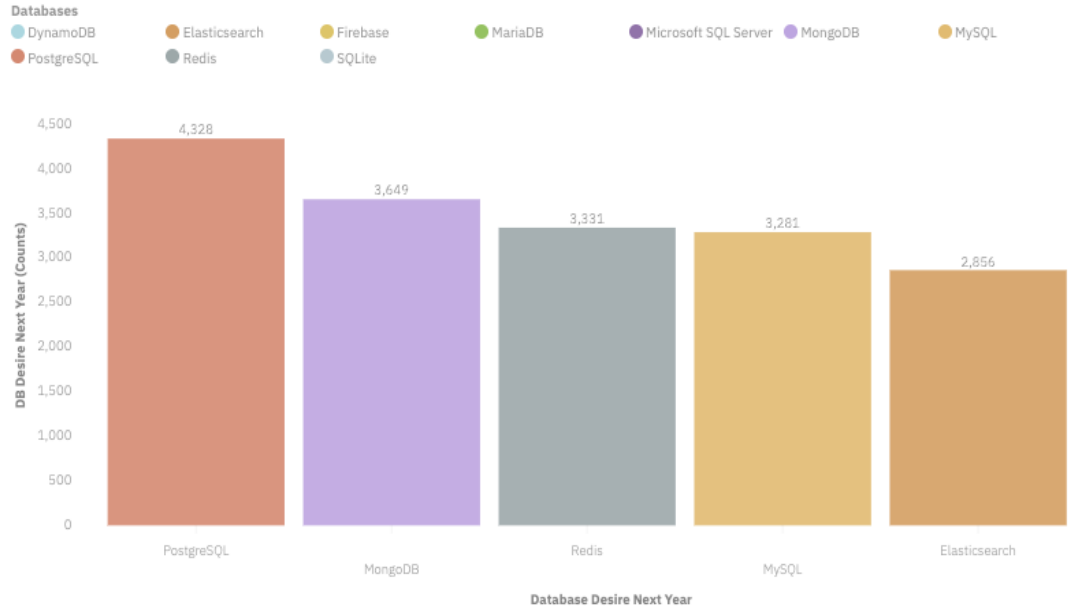
- Web Development still highly common among our respondents (**JavaScript** & **HTML**)
  - A follow-up to potentially query our JavaScript respondents on use for back-end systems and interface development with emerging blockchain technologies as potential points of interest
- **TypeScript** (a superset of JavaScript)
  - How many of our respondents above would similarly be layering on with their desire to learn TypeScript?
- Continuity of Data focused programmers using **Python**
  - Python's rise to third for future usage highlights its' significance in the evolving programming landscape and key role outside of more "Web" focused options above
- Similarly, **SQL** still a pivotal discipline for data storage/management
  - How we access and manage our data gathered appears to be ruled by SQL for foreseeable future
- How critical would lower interest in "**shell**" potentially be
  - Are "shell" scripters looking to migrate to newer technologies or is there less of a dependency on knowing them?

DATABASE TRENDS

Current Year



Next Year



Top 5 Databases

DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

* Below are a Few noteworthy findings

- Rise of **PostgreSQL** and **MongoDB**
  - Each db technology jumped multiple places in the top 5!
- **Microsoft SQL Server** with a high current usage but little future interest
  - Why such a drop-off for such a heavily used current DB technology?
- New Entries for **Elastic Search** & **Redis**
  - What are these new entries?
- **SQL** still a bedrock for utility of uses across different database platform (Ex. - PostgreSQL, MySQL)
  - No surprises here that SQL while not new like our two new entries is still the most common for current database technology usages.

Implications

* And potential implications

- Future Interest in **PostgreSQL** & **MongoDB**
  - Both a NoSQL (MongoDB) and SQL staple (PostgreSQL) indicate a desire for potential flexibility with each database design (Relational or Non-Relational).
- Dropoff in **Microsoft SQL Server**
  - Its' current high usage with less future interest than other databases could be a signaling of a shift in data management for future data storage
- **Elastic search** & **Redis**?
  - **Elastic search** - A fairly new distributed, free and open search db technology that helps with data ingestion from a variety of sources signals a potential new emerging platform for data management
  - **Redis –** Another NoSQL options for various data structures (strings, hashes, sets, ...etc) for in-memory distributed Key-Value storage
  - Looks like some new data storage options could disrupt **SQL's** dominance in the coming years!

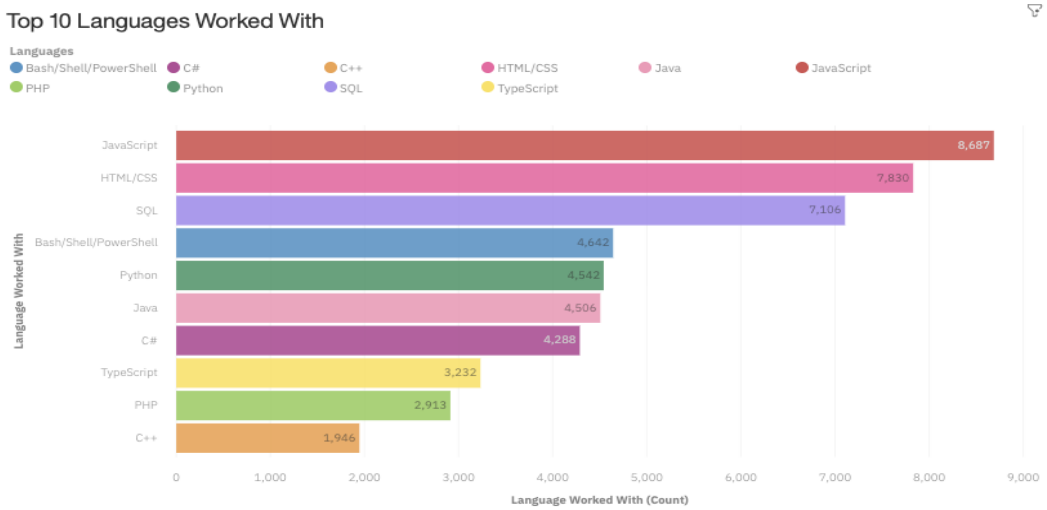
DASHBOARD

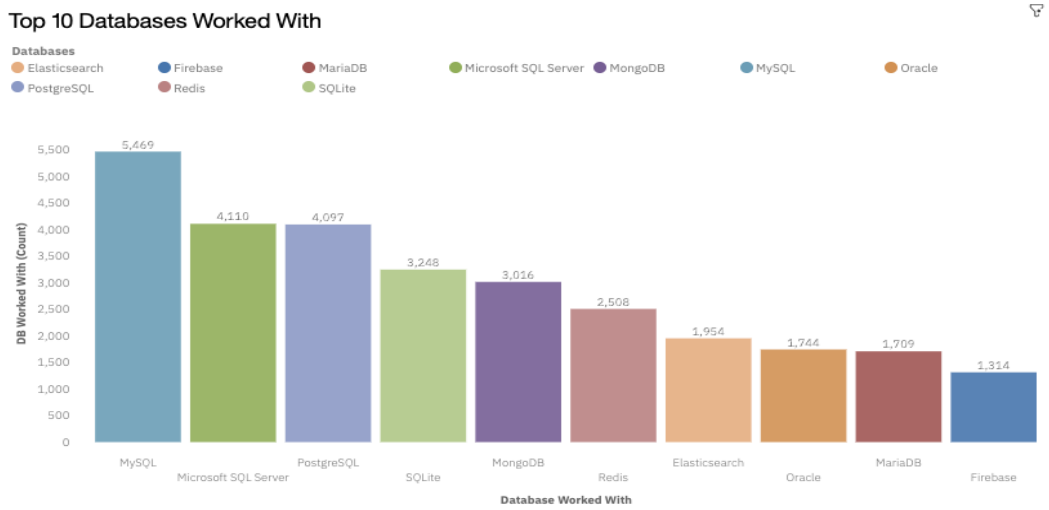
Read - Only Link to Cognos Dashboard

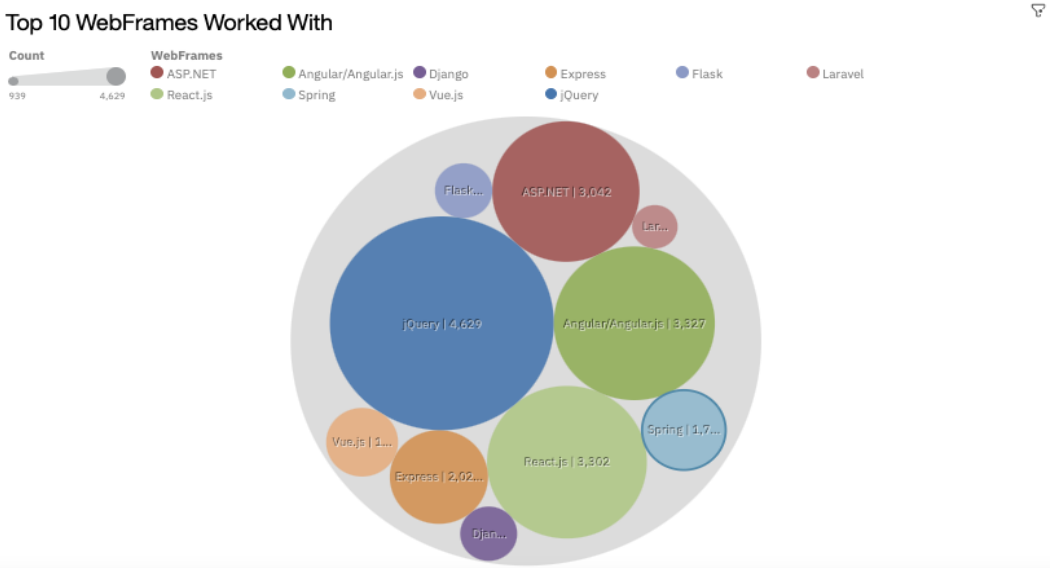
- Please see the link below as well
- https://dataplatform.cloud.ibm.com/dashboards/61d7561d-fb2d-4820-acd8-f05119dffc19/view/7c18ea2c0f9539e976d3b1e407cc2d502930215ae7bbd20ad6847b490f607097f06b1092c82f1e0fd2430665f6ee125ec0

DASHBOARD TAB 1

Current Technology Usage | Future Technology Trend | Demographics



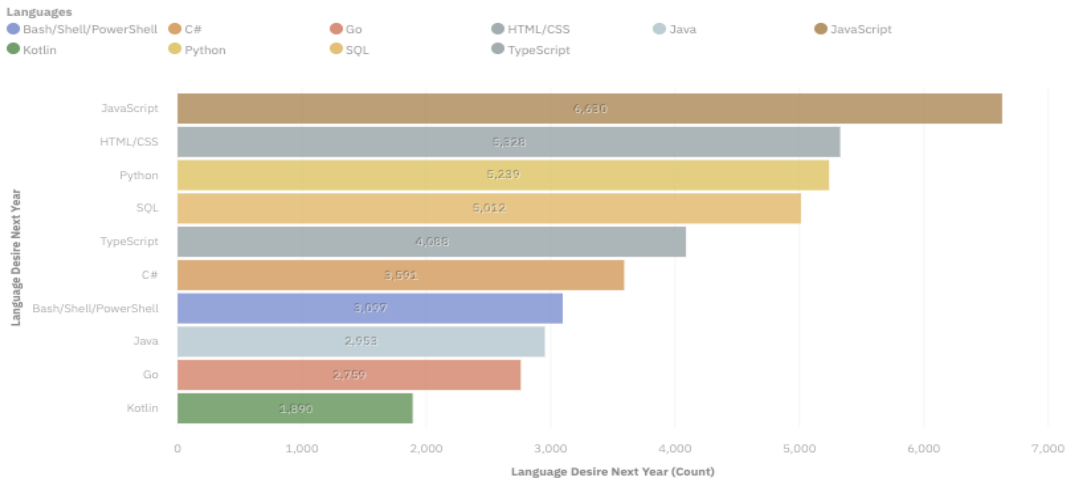




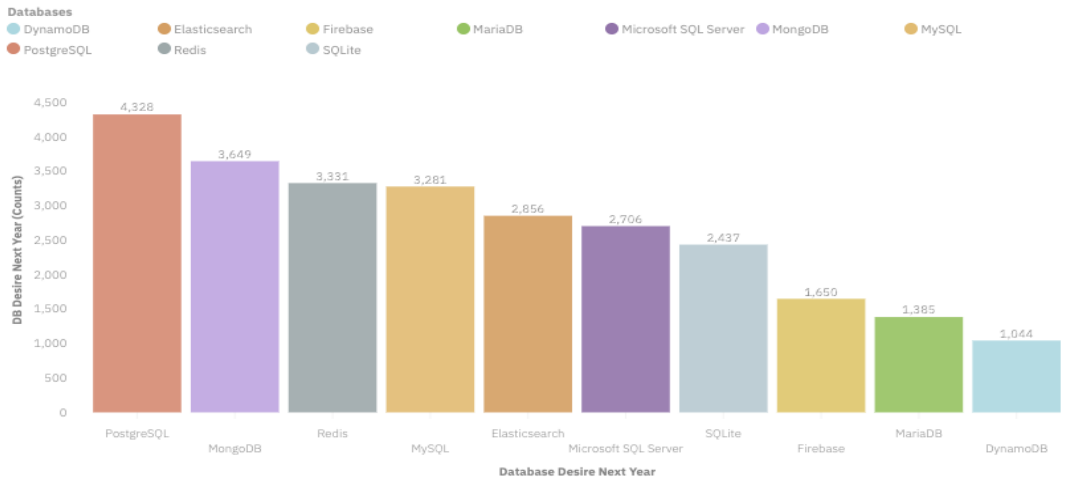
DASHBOARD TAB 2

Current Technology Usage | Future Technology Trend | Demographics

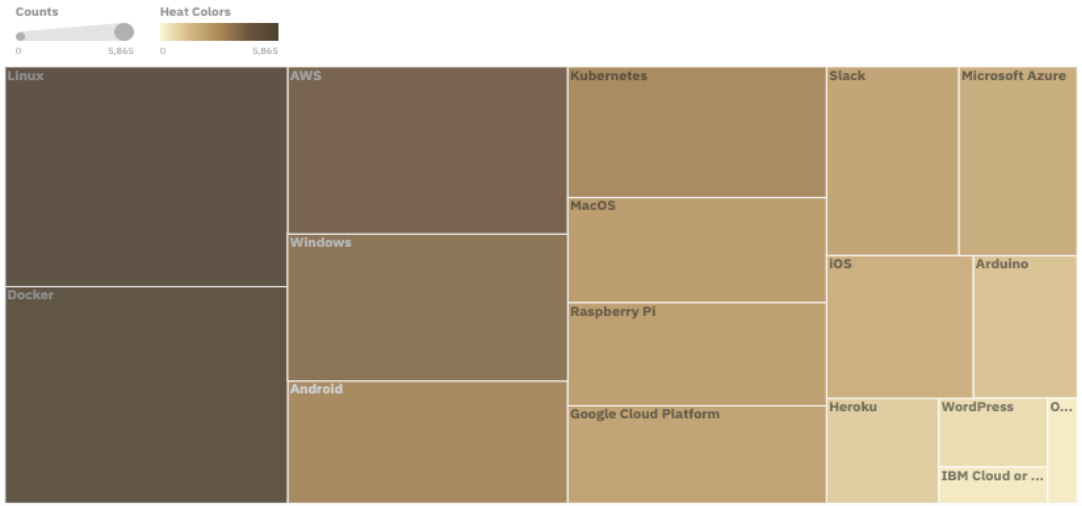
Top 10 Language Desire For Next Year



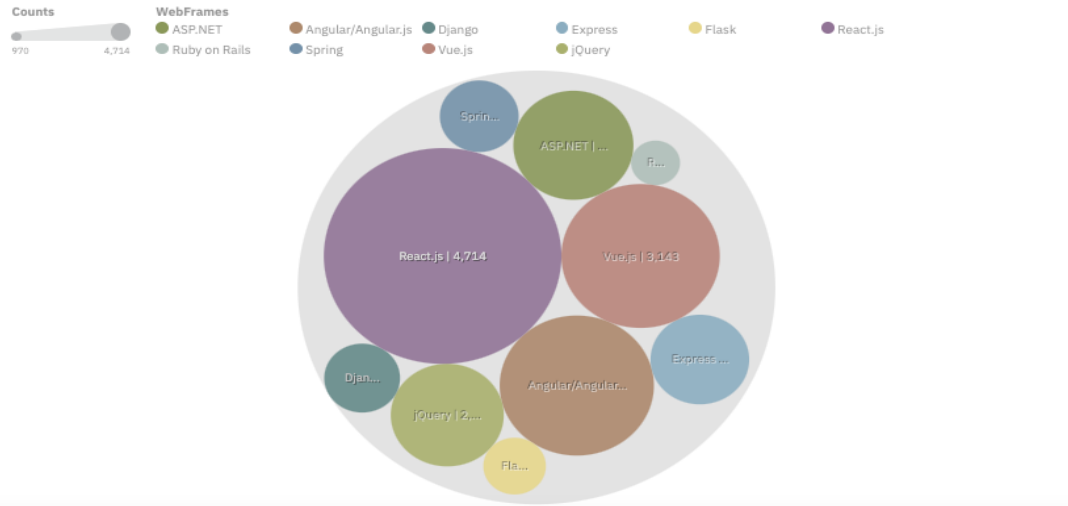
Top 10 DB Desire For Next Year



Platform(s) Desire For Next Year



Top 10 WebFrame Desire For Next Year



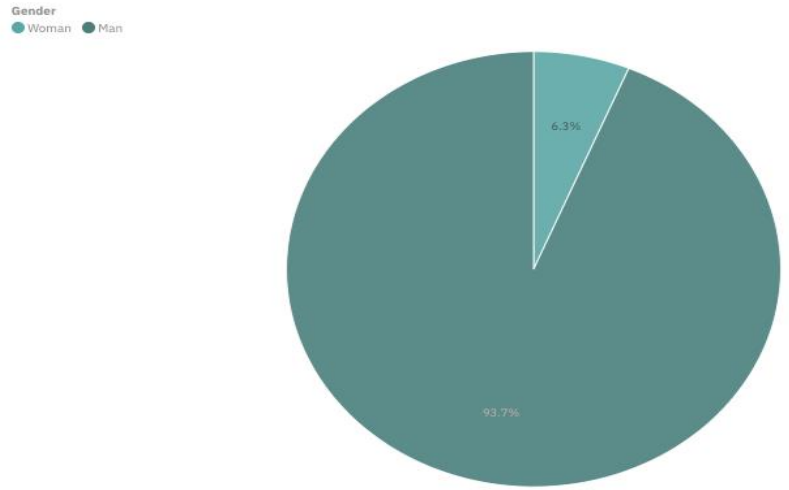
IBM Developer

SKILLS NETWORK

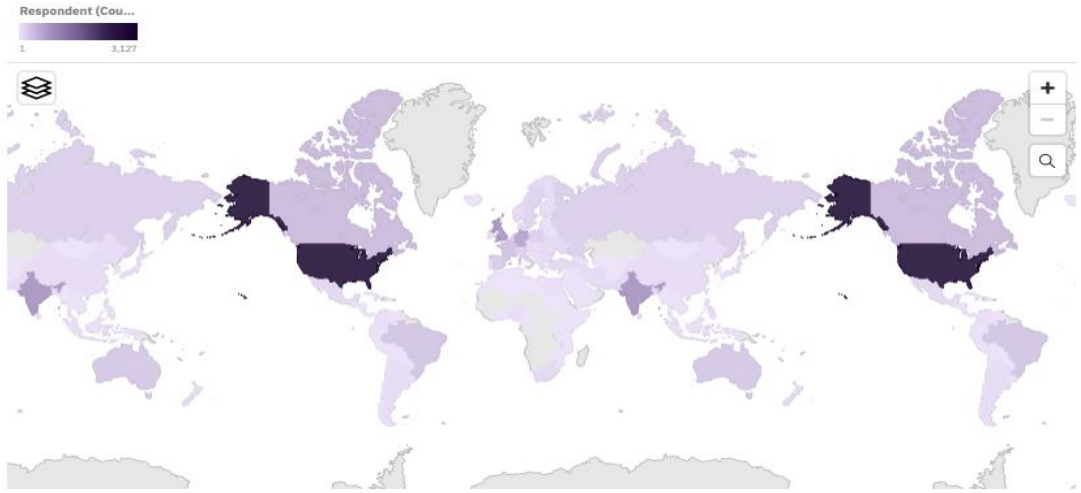
DASHBOARD TAB 3

Current Technology Usage | Future Technology Trend | Demographics

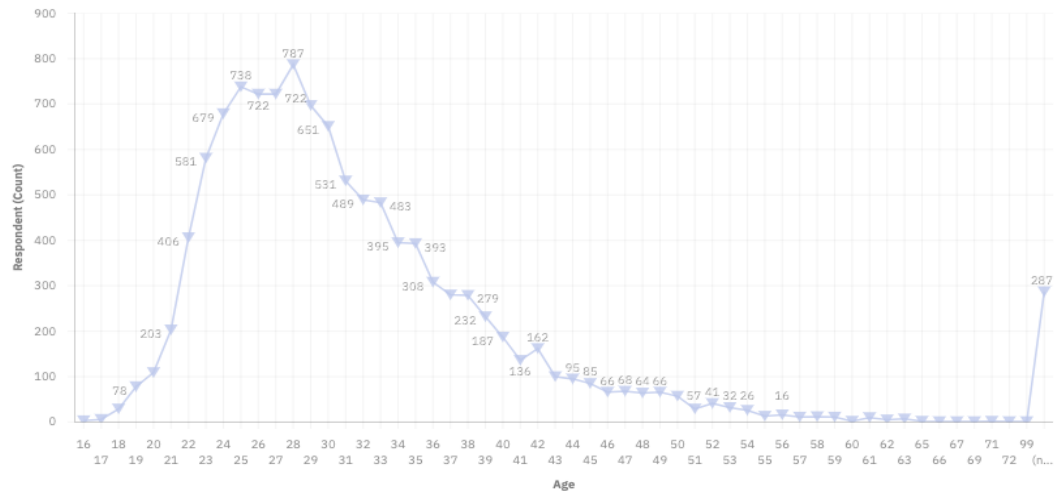
Respondent % Classified By Gender



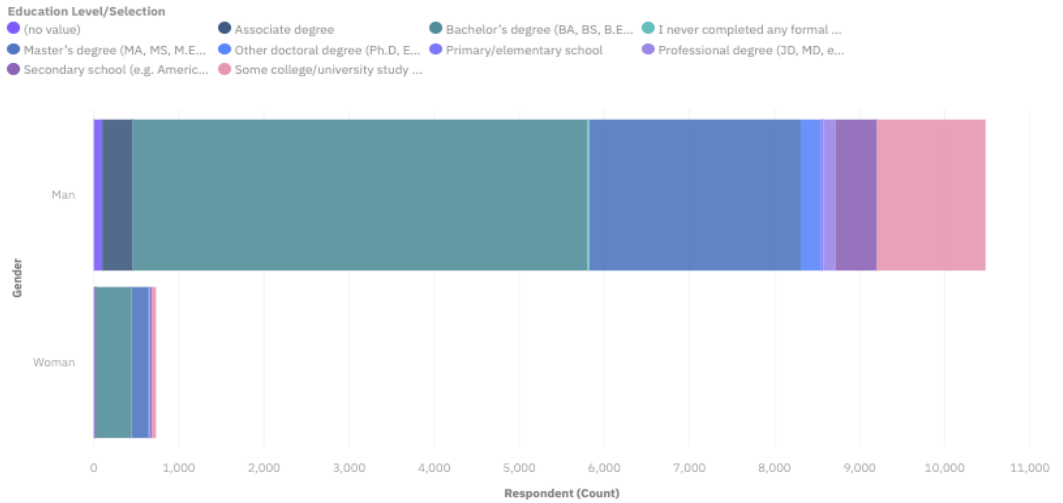
Respondent Count Per Country



Respondent Count By Age



Respondent Count Per Gender By Education Level



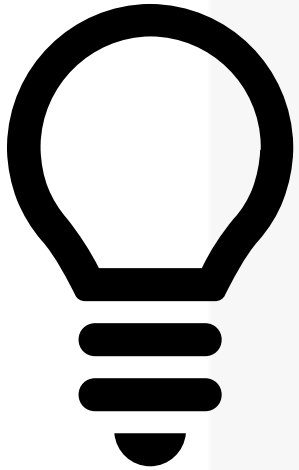
Discussion Topics (I of II)

- Programming & Framework Discussion Topics
  - Front-End Our Most Popular Respondents?
    - Seeing such a high respondent rate for front-end technologies and in particular the JavaScript frameworks, suggests potentially front-end developers may have been more willing to answer the survey or use stack overflow more commonly than other developers
  - Python Cementing in Second Place
    - Python appears to have solidified its' place behind JavaScript with the forefront of Data Analysis/Science and Machine Learning/AI potentially indicators of the future considerations for usage which appears to have more interest than such "shell" language usage or PHP/Java
  - Changing Tides
    - Alluded to above but it would appear that as interest rises in JavaScript and Python, interest is cooling for Java, PHP, Shell. The world surrounding programming is ever-changing and these numbers appear to indicate what's established now may not be what is to be most used moving forward.
- Database Discussion Topics
  - SQL Top Dog … for now
    - SQL has been and looks to continue to be the "godfather" of database usage. However, will changing data storage and more lightweight techniques continue to eat into future interest?
    - Tracking usage for Microsoft SQL Server, MySQL and SQLite will be interesting to see if continuing trends show an enthusiasm for newer technologies

Discussion Topics (II of II)

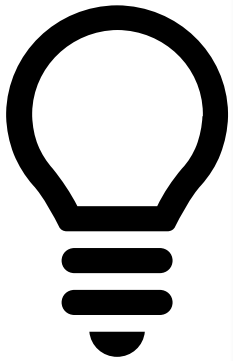
- Database Discussion Topics Cont.
  - Rise of NoSQL or NonSQL Storage
    - Redis and Elasticsearch highlight two evolving technologies for dynamic data storage and a keen interest in our respondents to continue to test and use its' capabilities
    - Can this continue to eat into SQL's dominance and make further impact for businesses moving forward who may be assessing "costs" and what is essential to hold in house for data storage?
- Demographics
  - Gender Disparity
    - With nearly 95% of our respondents being male, would a growing proportion of females in subsequent surveys influence our findings?
      - A poll later in this presentation does show that proportion losing some of its' dominance (about 92% male respondent rate in subsequent survey (see appendix))
  - Regional Bias Influence
    - A majority of our respondents are centered in a few locations. If our survey had higher respondent rates in other regions, would our results look similar?
  - Education Diversity
    - While a Batchelor's or Master's Degree is most common for our respondents, we can see that many education paths can end up finding a home in programming and database usage.
    - Pathways to programming & db usage appear quite diverse and help re-inforce the idea of their pervasiveness and effectiveness in helping industries far and wide.

CONCLUSION (I of IV)



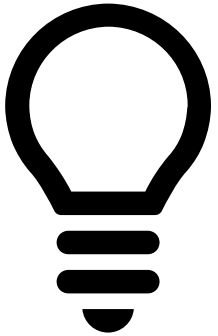
- Programming & Framework Takeaways
  - JavaScript Dominance
    - Whether used for programming, frameworks (**TypeScript**), or even with modern data storage with JSON key/value type lookup (**Redis**, **MongoDB**), it's safe to say JavaScript will continue to make its' mark on the industry.
    - Ranking the highest for individual programming language usage (and desired future usage) along with the highest for frameworks overall (**React**, **Vue**, **Angular**, and **Express**), the respondents polled indicate just how influential it is within our survey.
  - Growing Python Influence
    - Similar to JavaScript, the common pairing of **PostgreSQL** with Python Data Analysis and Programming is of keen interest and usage to our respondents.
    - **Flask** also shows it's influence within frameworks but unlike JS does not have nearly the array of frameworks of current and future users as mentioned above.
  - Changing Landscape
    - Programming languages such as **PHP, Java,** and **Shell Scripting** all in the top 10 for current usage, decrease significantly in future ranking or even out of the top 10 altogether.
    - Just as ever, as technology changes, we may continue to see new languages or frameworks surpass former technologies that had heavier influence in our respondents usage.

CONCLUSION (II of IV)



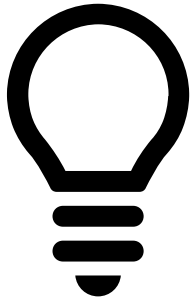
- Database Usage & Data Storage
  - Still a **SQL** World … yet a Changing Landscape
    - Current statistics don't lie, whether it's **PostgreSQL**, **Microsoft SQL Server,** or the "godfather" **MySQL (**Current Top Spot**); SQL** sets the agenda for a predominant proportion of current database usage.
    - **Microsoft SQL Server** appears to have a dominant usage but may be seen as an aging data storage option by falling from second in our current usage to sixth in our future visualization.
  - Diversity the path moving forward
    - With **PostgreSQL** and **MongoDB** being our top two database technologies moving forward, our respondents indicate that a **NoSQL** approach (**MongoDB**) and traditional **SQL** approach (**PostgreSQL**) is a trend to keep an eye on as data management continues to evolve over different disciplines (web/data) within programming.
    - **Elastic/Redis** illustrate a dynamic storage option for various data structures (strings, hashes, lists, sets, etc) that appear to be gaining traction and present new non-traditional **SQL** options for developers moving forward
  - Database Summary
    - Where **SQL** and specifically **PostgreSQL** appear to be the path for programmers today and moving forward, we likely will continue to see challengers outside of **SQL** for more flexible database storage than traditional relational **SQL** models.

CONCLUSION (III of IV)



- Demographic Review
  - US, UK/Europe, & India Lead the Way for Survey Respondents
    - From the geographic map we can see the US leads the way with the other mentioned countries/regions just behind.
    - Brazil, Australia, and Russia also have had an impact on our results as countries making up a majority of respondents after our top regions/countries above.
  - Starting Young
    - We see an exponential growth in our respondents' age in their very early 20s. (21 is the first major upswing)
    - The age consolidates and generates an interquartile range and highest density near the ages 22-35.
    - Our data may be slightly biased as Stack Overflow has become a popular resource within the past decade.
  - Heavily skewed toward male respondents
    - Males make up nearly 95% of all respondents.
    - A global developer survey in 2022 found a similar proportion with nearly 92% of all respondents to be males.
      - https://www.statista.com/statistics/1126823/worldwide-developer-gender/
    - The study above does show a decreasing proportion from males from our data (2019) to the above survey respondents. A moving trend showing more females potentially in subsequent surveys would be of interest to track.

CONCLUSION (IV of IV)

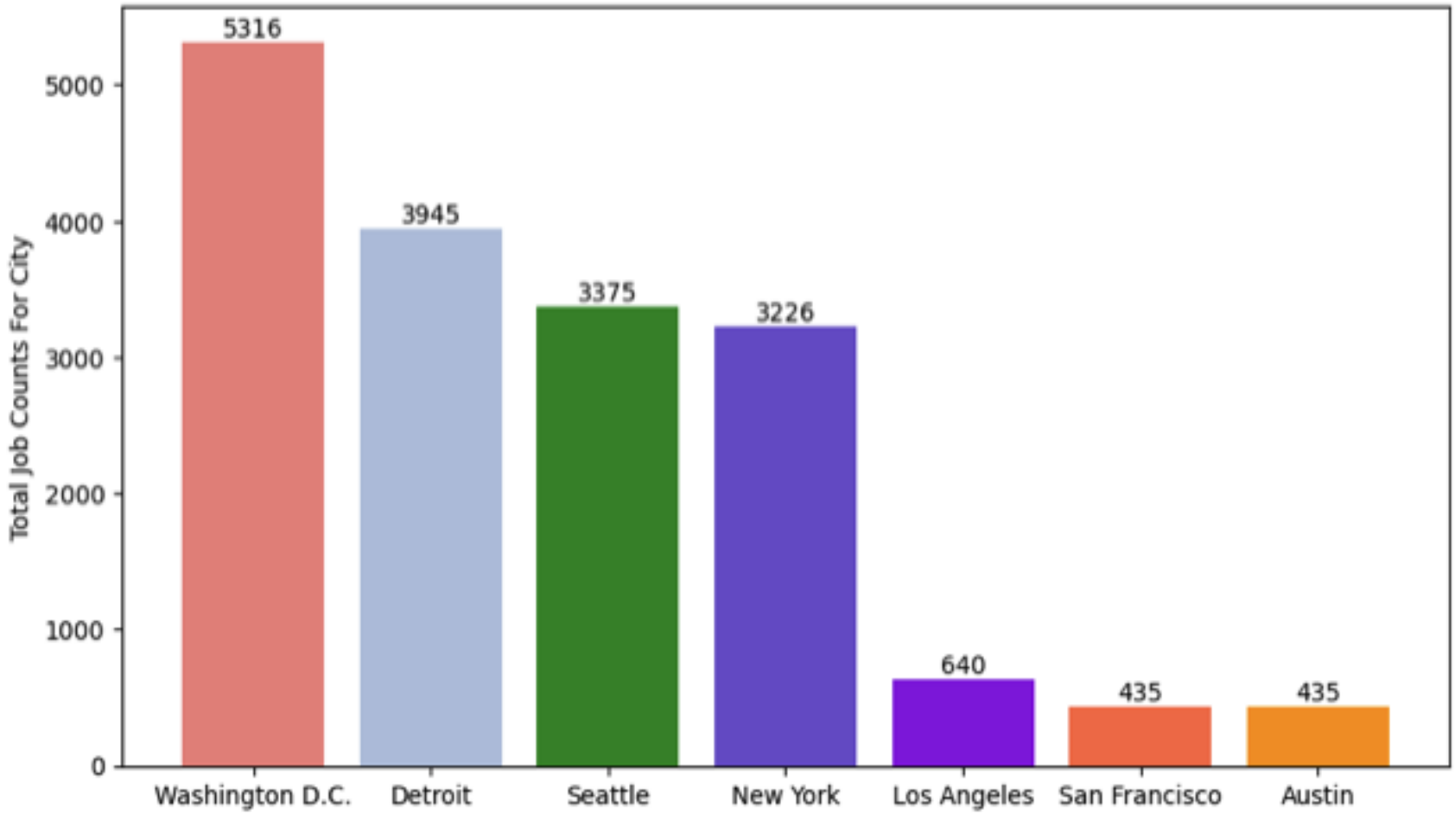


- Demographics Continued
  - Varying Education But Generally Some College
    - For males and females the most dense education selection was a bachelor's degree followed by a master's degree.
    - Either some college or a professional degree can also be seen to have an impact which would suggest alternative education paths as a route for our respondents. However, a bit surprising to see such professional degrees like either a JD or MD also having a notable impact in our proportional make-up.
  - No One Way … yet a common pattern of some education generally
    - A bit of a cliché but our various education category selections do indicate that there is a formal "college" element with a majority of the respondents, although that there are alternative education paths.
    - Only a small proportion shows no type of education but there is enough of a chunk to show that completion of a degree isn't the only way
  - Quick Demographic Summary
    - Male dominant in terms of proportional responses but reason to suspect that could decrease based off a more recent survey showing a lower percentage than our gender proportions.
    - Some to various types of formal education highlight programming's pervasive nature into varying industries as well as our respondents likely diverse educational make-up outside of just software engineering.
    - Our respondents' age does highlight that a majority of the respondents appear to be early to mid-professional within their careers with a very "right skewed" age distribution.

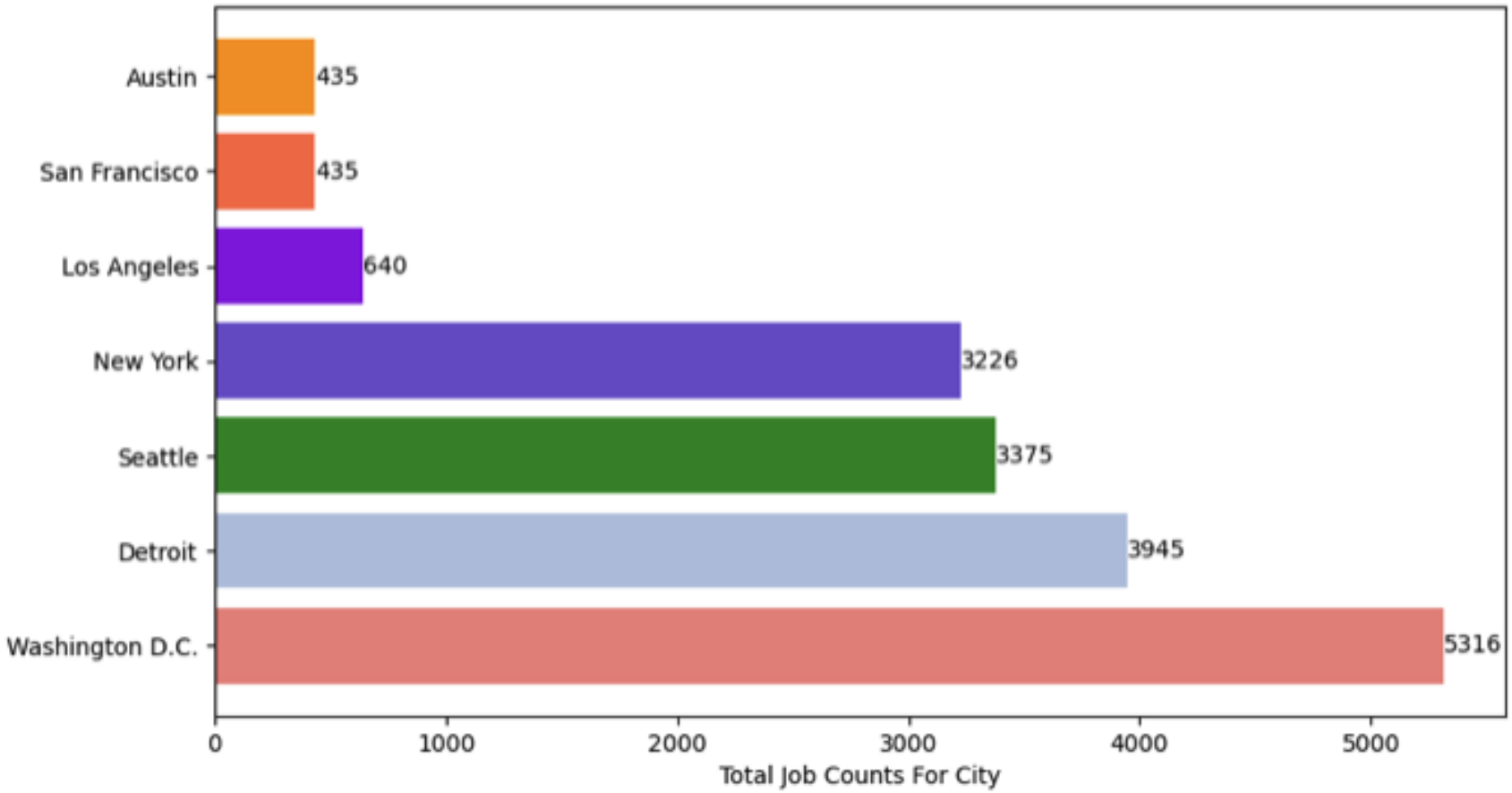
Data Visualizations (See Next Slides)

- Bar Graphs for Template Rubric Prompts
  - Job Counts Per API Return for Active Job Listings for Available Cities
  - Median Salary for Popular Programming Languages

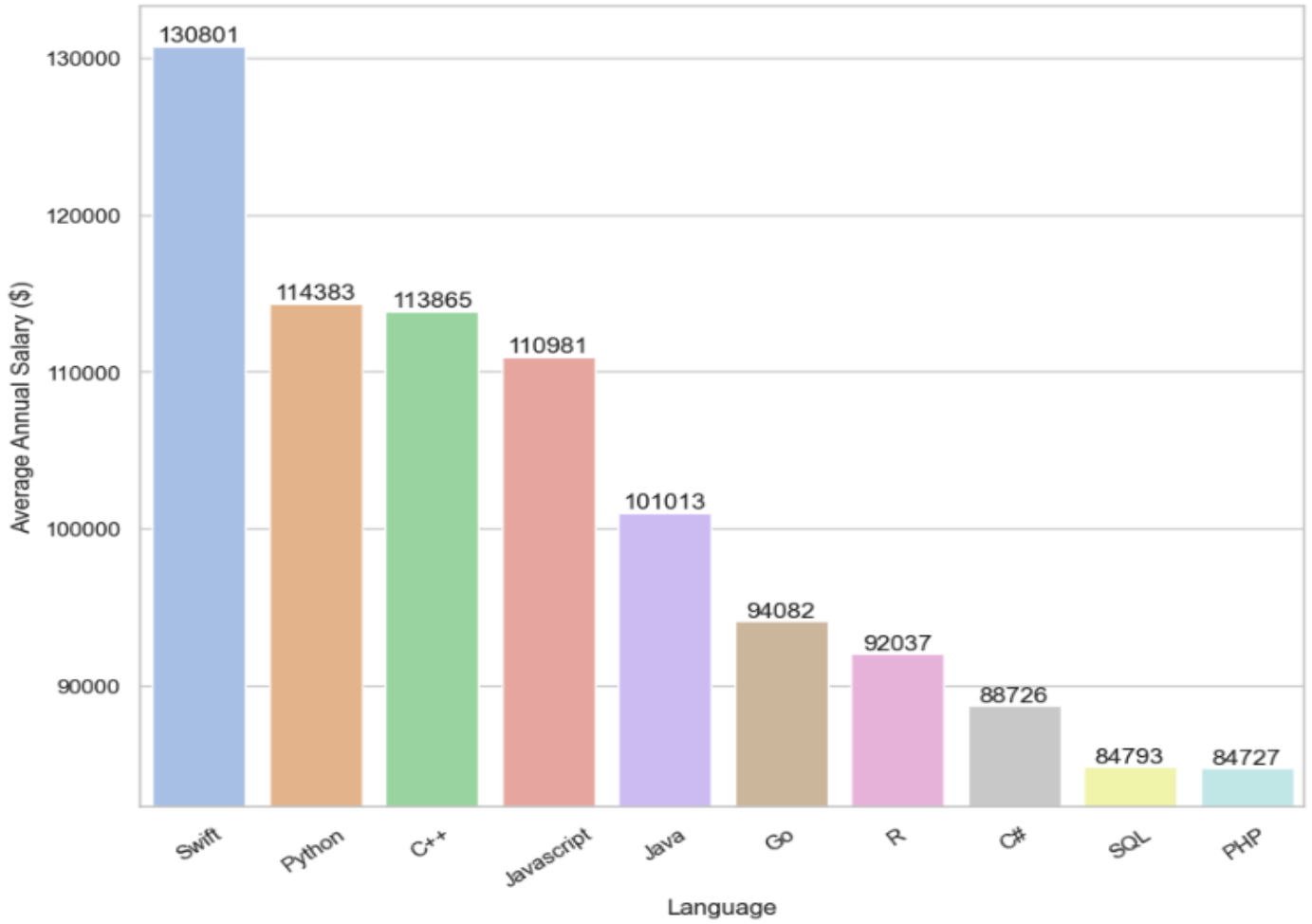
Job Counts for Cities from Job API Returns



Job Counts for Cities from Job API Returns



Popular Languages Average Salary (USD)



APPENDIX

- Data Sources
  - Stack Overflow Programming Data pulled from 2019 Report
    - https://stackoverflow.blog/2019/04/09/the-2019-stack-overflow-developer-survey-results-are-in
  - Job Posting Data
    - https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DA0321EN-SkillsNetwork/labs/module%201/Accessing%20Data%20Using%20APIs/Jobs_API.ipynb
  - Programming Language Salaries
    - https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DA0321EN-SkillsNetwork/labs/datasets/Programming_Languages.html